Cell, Vol. 50, 667, August 28, 1987, Copyright © 1987 by Cell Press

Letter to the Editor

"Homology" in Proteins and Nucleic Acids: A Terminology Muddle and a Way out of It

"Homology" has the precise meaning in biology of "having a common evolutionary origin," but it also carries the loose meaning of "possessing similarity or being matched." Its rampant use in the loose sense is clogging the literature on protein and nucleic acid sequence comparisons with muddy writing and, in some cases, muddy thinking.

In its precise biological meaning, "homology" is a concept of quality. The word asserts a type of relationship between two or more things. Thus, amino acid or nucleotide sequences are either homologous or they are not. They cannot exhibit a particular "level of homology" or "percent homology." Instead, two sequences possess a certain level of *similarity*. Similarity is thus a quantitative property. Homologous proteins or nucleic acid segments can range from highly similar to not recognizably similar (where similarity has disappeared through divergent evolution).

If using "homology" loosely did not interfere with our thinking about evolutionary relationships, the way in which we use the term would be a rather unimportant semantic issue. The fact is, however, that loose usage in sequence comparison papers often makes it difficult to know the author's intent and can lead to confusion for the reader (and even for the author).

There are three common situations in which hazards arise by using "homology" to mean similarity. The first case is the most obvious offense but perhaps the least troublesome. Here an author identifies sequence similarities (calling them homologies) but claims that the sequences being compared are not evolutionarily related. Some awkward moments occur in such a paper, since the author claims both homology (i.e., similarity) and nonhomology (i.e., lack of a common ancestor). Nonetheless, the author's ideas are likely to be clear since arguments against common ancestry are presented explicitly.

A second case is one in which an author points out similarities (again called homologies) but does not address the issue of evolutionary origins. The reader, seeing the term "homology," may infer that the author is postulating coancestry when that is not the author's intent.

The final case occurs most frequently and is the most subtle and therefore most troublesome. Here, similarities (called homologies) are used to support a hypothesis of evolutionary homology. In this case, the two meanings of homology seem to overlap, and it is almost inevitable that the thinking of author and reader alike will be intrusively distorted as follows. Similarity is relatively straightforward to document. In comparing sequences, a similarity can take the form of a numerical score (% amino acid or nucleotide positional identity, in the simplest approach) or of a probability associated with such a score. In comparisons of three-dimensional structures, a typical numerical description is root-mean-square positional deviation between compared atomic positions. A similarity, then, can become a fully documented, simple fact. On the other hand, a common evolutionary origin must usually remain a hypothesis, supported by a set of arguments that might include sequence or three-dimensional similarity. Not all similarity connotes homology but that can be easily overlooked if similarities are called homologies. Thus, in this third case, we can deceive ourselves into thinking we have proved something substantial (evolutionary homology) when, in actuality, we have merely established a simple fact (a similarity, mislabeled as homology). Homology among similar structures is a hypothesis that may be correct or mistaken, but a similarity itself is a fact, however it is interpreted.

We believe that the concepts of evolutionary homology and sequence or three-dimensional similarity can be kept distinct only if they are referred to with different words. We therefore offer the following recommendations:

- Sequence similarities (or other types of similarity) should simply be called similarities. They should be documented by appropriate statistical analysis. In writing about sequence similarities the following sorts of terms might be used: a level or degree of similarity; an alignment with optimized similarity; the % positional identity in an alignment; the probability associated with an alignment.
- Homology should mean "possessing a common evolutionary origin" and in the vast majority of reports should have no other meaning. Evidence for evolutionary homology should be explicitly laid out, making it clear that the proposed relationship is based on the level of observed similarity, the statistical significance of the similarity, and possibly other lines of reasoning.

One could argue that the meaning of the term "homology" is itself evolving. But if that evolution is toward vagueness and if it results in making our scientific discourse unclear, surely we should intervene. With a collective decision to mend our ways, proper usage would soon become fashionable and therefore easy. We believe that we and our scientific heirs would benefit significantly.

Gerald R. Reeck,[1] Christoph de Haën,[2] David C. Teller,[3] Russell F. Doolittle,[4] Walter M. Fitch,[5] Richard E. Dickerson,[6] Pierre Chambon,[7] Andrew D. McLachlan,[8] Emanuel Margoliash,[9] Thomas H. Jukes,[10] and Emile Zuckerkandl[11]

[1]Department of Biochemistry, Kansas State University, Manhattan, Kansas 66506; [2]Departments of Medicine and Biochemistry, University of Washington, Seattle, Washington 98195; [3]Department of Biochemistry, University of Washington, Seattle, Washington 98195; [4]Department of Chemistry, University of California at San Diego, La Jolla, California 92093; [5]Department of Molecular Biology, University of Southern California, Los Angeles, California 90089; [6]Molecular Biology Institute, University of California, Los Angeles, California 90024; [7]Institute of Biological Chemistry, Strasbourg, France; [8]MRC Laboratory of Molecular Biology, Cambridge, England; [9]Department of Biochemistry and Molecular Biology, Northwestern University, Evanston, Illinois 60201; [10]Department of Biophysics, University of California, Berkeley, California 94720; [11]Linus Pauling Institute, Palo Alto, California 94306

EXHBIIT B

48. D. W. Salter, E. J. Smith, S. H. Hughes, S. E. Wright, L. B. Crittenden, *Virology* **157**, 236 (1987).
49. R. A. Bosselman *et al.*, *Science* **243**, 533 (1989).
50. D. W. Salter and L. B. Crittenden, *Theor. Appl. Genet.*, in press.
51. K. Roshlau *et al.*, *Arch. Tierz.* **31**, 3 (1988).
52. Z. Zhu, K. Xu, G. Li, Y. Xie, L. He, *Kexue Tongbao* **31**, 988 (1986).
53. K. Ozato *et al.*, *Cell Differ.* **19**, 237 (1986).
54. D. Chourrout, R. Guyomard, L.-M. Houdebine, *Aquaculture* **51**, 143 (1986).
55. R. A. Dunham, J. Eash, J. Askins, T. M. Townes, *Trans. Am. Fish. Soc.* **116**, 87 (1987).
56. G. Brem, B. Brenig, G. Hoerstgen-Schwark, E.-L. Winnacker, *Aquaculture* **68**, 209 (1988).
57. T. McEvoy *et al.*, *ibid.*, p. 27.
58. G. W. Stuart, J. V. McMurray, M. Westerfield, *Development* **103**, 402 (1988).
59. E. J. C. Polge *et al.*, in *Biotechnology of Growth Regulation*, R. B. Heap *et al.*, Eds. (Butterworth, London, in press).
60. K. L. Knight, H. Spieker-Polet, D. S. Kaydin, V. T. Oi, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 3130 (1988).
61. J. P. Simons *et al.*, *Bio/Technology* **6**, 179 (1988).
62. C. E. Rexroad, Jr., *et al.*, *Mol. Reprod. Dev.* **1**, 164 (1989).
63. J. D. Murray, C. D. Nancarrow, J. T. Marshall, I. G. Hagelton, K. A. Ward, *Reprod. Fertil. Dev.*, in press.
64. R. L. Brinster, in *Growth, Nutrition and Metabolism of Cells in Culture*, G. Rothblat and V. Cristofalo, Eds. (Academic Press, New York, 1972), pp. 251–286.
65. From D. J. Bolt, U.S. Department of Agriculture, Beltsville, MD.
66. D. N. Marple and E. D. Aberle, *J. Anim. Sci.* **34**, 261 (1972).
67. L. A. Frohman and T. R. Downs, *Methods Enzymol.* **124**, 371 (1986).
68. R. W. Furlanetto *et al.*, *J. Clin. Invest.* **60**, 648 (1977).
69. W. H. Daughaday, I. K. Mariz, S. L. Blethen, *J. Clin. Endocrinol. Metab.* **51**, 781 (1980).
70. Gastric ulceration: focal loss of gastric mucosa with local inflammation in the pars esophagus and variable evidence of granulation tissue in healing lesions. Synovitis: inflammation of joint synovium, ranging from mild edema with focal aggregates of inflammatory cells to necrosis and ulceration of thickened synovium with erosion of underlying cartilage. Dermatitis: focal perivascular accumulation of inflammatory cells in subcutis and superficial dermis and mild epidermal thickening. Nephritis: mild, multifocal accumulation of lymphocytes in tubulointerstitial space. Pneumonia: inflammation of lung, ranging from mild interstitial accumulation of lymphocytes to multifocal perivascular and peribronchiolar accumulation of lymphocytes and plasma cells consistent with enzootic pneumonia. Cardiac myocyte nuclear hypertrophy: enlargement of numerous cardiac myocyte nuclei of unknown significance.
71. We thank L. L. Schulmann, M. K. Chandlee, N. J. Faulkner, M. Avarbock, L. Baker, and M. K. Yagle for technical assistance and K. D. Bender, T. E. Clover, J. A. Piatt, and J. E. Stevens for animal care. We acknowledge the assistance of K. E. Mayo and L. A. Frohman in analyzing the GRF pigs; T. J. Van Winkle and D. M. Gillette in performing the necropsies and the histologic work-up; and A. D. Mitchell and M. B. Solomon in evaluating growth performance and carcass composition. We thank R. R. Behringer, C. E. Rexroad, Jr., E. P. Sangren, and N. C. Steele for assistance and helpful discussion in preparing the manuscript. Supported by NIH research grants (HD-19018) to R.L.B. and (HD-09172) to R.D.P.

Altering the Genome by Homologous Recombination

MARIO R. CAPECCHI

Homologous recombination between DNA sequences residing in the chromosome and newly introduced, cloned DNA sequences (gene targeting) allows the transfer of any modification of the cloned gene into the genome of a living cell. This article discusses the current status of gene targeting with particular emphasis on germ line modification of the mouse genome, and describes the different methods so far employed to identify those rare embryonic stem cells in which the desired targeting event has occurred.

THE IMPLICATIONS OF THE NEW GENE TARGETING TECHnology are far-reaching. If the recipient cell is a pluripotent, embryo-derived stem (ES) cell, it is possible to transfer a modification of a cloned gene, created in a test tube, to the germ line of a living organism (*1–3*). The potential now exists for modifying any gene, in a defined manner, in any species from which functional ES cells can be obtained. ES cells have been isolated from mouse and hamster embryos (*4*) and major efforts are currently under way to isolate equivalent cells from domestic animals including sheep, pigs, and cattle. In addition, because many plant cells are intrinsically pluripotent and the means exist for generating whole plants from these cultured cells, we can anticipate the application of gene targeting to the modification of plant genomes as well (*5*).

The discussion need not, however, be limited to experiments directed only at germ line modifications. In specific cases it may be advantageous to modify only certain somatic tissues of an organism. For example, as the means to propagate a variety of human somatic stem cells (such as hematopoietic, epithelial, liver, or lung stem cells) become available, protocols based on gene targeting could be used to correct defective genes in the appropriate human tissue. This scenario of human somatic gene therapy has some obvious advantages over the random insertion of a nondefective gene: for example, the corrected endogenous gene is much more likely to be expressed in the appropriate tissue at appropriate levels. Further, it should be possible to use this approach to correct dominant mutations.

In addition to its implication for in vivo manipulations, gene targeting technology has broad potential for fundamental research with cells cultured in vitro. Many biological questions can be answered directly and more simply with tissue culture systems. In such cells, both alleles of an autosomal gene could be modified by the sequential application of gene targeting. Cell-lethal phenotypes could be maintained and analyzed by a variety of techniques, including the introduction of a transgene under the control of an inducible promoter.

However, in this article I will emphasize experiments involving mouse embryo-derived stem cells. This choice is based on the interest and potential of using targeted, modified ES cells as a vehicle to generate mice of any desired genotype. Unfortunately, this choice precludes reviewing the gene targeting literature leading

The author is associated with the Howard Hughes Medical Institute, Department of Biology, University of Utah, Salt Lake City, UT 84112.

up to these experiments. For those interested in this literature, I would recommend the 1984 Cold Spring Harbor Symposium on homologous recombination (*6*) and a recent review (*7*). I will also discuss various approaches being used to modify nonselectable genes and venture some guesses as to where the field is likely to progress.

From ES Cells to Germ Line Chimera

Figure 1 outlines the procedure for introducing a designed mutation into the germ line of mice by means of targeted modification of the ES cell genome. A targeting vector, containing the desired mutation, is introduced into ES cells by electroporation or microinjection. In most cells the targeting vector inserts randomly into the ES genome. However, in a few cells, the targeting vector pairs with the cognate chromosomal DNA sequence and transfers the mutation to the genome by homologous recombination. Screening or enrichment procedures (or both) are then used to identify the rare ES cell in which the targeted event has occurred. The appropriate cell is then cloned and maintained as a pure population. Next, the altered ES cells are injected into the blastocoel cavity of a preimplantation mouse embryo and the blastocyst is surgically transferred into the uterus of a foster mother where development is allowed to progress to term. The resulting animal is chimeric in that it is composed of cells derived from both the donor stem cells and the host blastocyst. In the particular example shown in Fig. 1, the ES cells are derived from a mouse homozygous for the black coat color allele and the recipient blastocyst is derived from an albino mouse. The fur of the resulting chimeric mouse has patches of both colors because the mouse contains cells of both genotypes. Breeding of the chimeric mouse to an albino mouse yields some black mice, indicating that the ES cells contributed to the formation of the germ line. Genomic screening of these progeny is used to determine which mice received the allele carrying the targeted mutation. Interbreeding of heterozygous siblings yields animals homozygous for the desired mutation.

Disruption of *hprt*

Mammalian cells can mediate recombination between homologous DNA sequences but they demonstrate an even greater disposition for mediating nonhomologous recombination. The problem is thus to identify homologous recombination events in a vast pool of scattered, nonhomologous recombination events. The hypoxanthine phosphoribosyl transferase gene (*hprt*) has provided an ideal model system for developing the technique of gene targeting in ES cells, because one can select directly for the targeting event. Since this gene is on the X chromosome, only one mutant copy is needed to yield the recessive *hprt*$^-$ phenotype in male ES cells. The *hprt*$^-$ cells are selected by growth in the presence of the base analog, 6 thioguanine (6-TG), which kills *hprt*+ cells.

Figure 2 illustrates the use of sequence replacement and sequence insertion vectors to disrupt *hprt* (*8*). Using yeast as a paradigm, we anticipated that sequence replacement vectors would replace endogenous DNA with exogenous sequences, whereas sequence insertion vectors would insert the entire vector DNA sequence into the endogenous locus. Since the final products are predicted to be different when these two classes of vectors are used (note the partial duplication of the gene in Fig. 2B), each vector could generate different types of mutant alleles. The presence of the gene encoding neomycin phosphotransferase (*neo*) within the eighth exon of *hprt* disrupts the *hprt* coding sequence and also provides a selectable

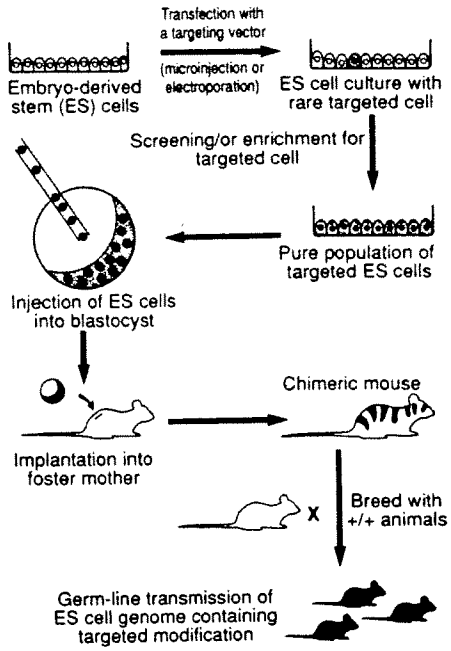

Fig. 1. Generation of mouse germ line chimeras from embryo-derived stem (ES) cells containing a targeted gene disruption.

marker (resistance to the drug G418).

Both types of targeting vector were introduced into ES cells by electroporation. The transfectants that survived selection in G418 and 6-TG had lost *hprt* activity as a result of a targeted disruption of *hprt*. Both types of vectors were equally efficient at disrupting the endogenous *hprt* (*8*). Furthermore, replacement and insertion vectors showed the same dependency of the targeting frequency on the extent of homology between the targeting vector and endogenous DNA sequences (Fig. 3). (The word "homology" is used here to describe participants in homologous recombination, which are generally identical.) Over the range tested, from 2.9 to 14.3 kb, a fivefold increase in DNA sequence homology resulted in roughly a 100-fold increase in the targeting frequency. In the above experiments, the amount of nonhomology (*neo*) being transferred to the target was kept constant. Therefore, it has not been determined whether the critical parameter is the absolute extent of sequence homology between the incoming DNA and the target, or whether the relative amount of nonhomology is also important. However, with the largest targeting vector the absolute targeting frequency was one independent targeting event per 3×10^4 ES cells electroporated.

Smithies and his colleagues have corrected a defective *hprt* in ES cells using a sequence insertion vector (*9*) and have inactivated *hprt* with a sequence replacement vector (*10*). In the former experiments the recipient ES cells contained a spontaneous 5′ deletion in *hprt* and the incoming targeting vector supplied the missing exons. In the latter experiments the replacement-type vector contained only 1.3 kb of *hprt* sequence homology disrupted by a promoterless *neo*. As will be discussed shortly, use of a promoterless *neo* in the targeting vector yields an enrichment for homologous versus nonhomologous recombination events. More recently Thompson *et al.* (*3*) generated germ line chimeras from ES cells in which a mutant *hprt* had been corrected by gene targeting. This demonstrates that ES cells transfected by electroporation and subjected to HAT selection still retain their ability to contribute to a functional germ line.

It has not been straightforward to make quantitative comparisons among *hprt* targeting experiments done in different laboratories. Targeting vectors containing different extents of homology to the endogenous target were used. In addition, the targeting vectors have been directed to different regions of *hprt*. The simplest means for

comparison is in terms of absolute frequency of homologous recombination. How many cells were electroporated and how many independent targeting events were recovered? With this criterion for comparison and with allowance for the differing extents of homology to the target, the results from the different laboratories are in moderate agreement [When targeting vectors containing approximately 2 to 4 kb of *hprt* homology were used, absolute targeting frequencies in the range of one event per 5×10^7 to 5×10^6 treated cells were reported (*3, 8–10*)]. Alternatively, it is possible to compare the ratio of homologous recombination events to nonhomologous recombination events. Unfortunately this ratio is dependent on the method used to estimate the frequency of nonhomologous recombination. As an extreme example, if the comparison was made to the transfection efficiency measured with a completely defective selectable gene, this ratio would be infinitely large. Indeed a wide range of ratios have been reported, from 0.001% to greater than 10%. It should also be pointed out that even when the same selectable gene is used, in the same nucleotide environment, measurements of both the targeting and random insertion frequency may be less than ideal, since expression of the selectable gene will inevitably be different at the target locus relative to random loci. Thus, the ratio will be dependent on the thresholds set by the selection conditions.

Nonselectable Genes

The advantages of *hprt*—its presence as a single copy in male cells and the ability to use direct selection as a way of isolating homologous recombinants—are not the case for most genes of interest. Indirect enrichment or screening procedures must be used to identify the rare ES cell in which a nonselectable gene has been inactivated.

A very sensitive screening method makes use of the polymerase chain reaction (PCR) (*11*) to specifically amplify a novel DNA junction created by the targeting event (*12*). Two groups have recently reported success using this approach to screen pools of transfectants and detect the rare ES cell in which the targeting event occurred. Joyner *et al.* have disrupted the homeobox-containing gene *en-2* in ES cells by means of a replacement vector in which one of the *en-2* exons was interrupted by *neo* (*13*). After introducing the *en-2/neo*^{r} targeting vector into ES cells by electroporation and selecting for G418^r cells, the authors screened pools of G418^r colonies by PCR to specifically amplify the *neo*-disrupted *en-2*. They detected approximately one targeting event per 300 G418^r colonies.

Zimmer and Gruss have disrupted another mouse homeobox-containing gene, *hox 1.1* (*14*). They used a replacement vector in which the homeobox domain carried a 20-bp insert, which disrupted the coding sequence and created a novel hybridization site for one of the PCR primers. The targeting vector was introduced into ES cells by microinjection. Since no selectable marker was used, ES cells in which a targeting event had occurred were identified solely by the presence of a specific PCR product. Only homologous recombination would juxtapose the two PCR primers, one within the targeting vector and the other from flanking *hox 1.1* sequences in the endogenous gene. Approximately 1 in 150 cells receiving an injection yielded the predicted amplified PCR fragment.

It is also possible to use the cis-acting regulatory sequences of the target gene in order to enrich for successful recombination events. In this case, the targeting vector is designed in such a way that expression of the selectable marker, *neo*, depends on homologous integration to supply a missing promoter or enhancer, as described by Jasin and Berg (*15*) as well as Sedivy and Sharp (*16*). In this way, a several hundredfold enrichment for targeted integrations relative to random insertion of the vector was obtained. This strategy only applies to genes that are expressed in the recipient cell line.

The Positive-Negative Selection (PNS) Procedure

Recently, we have described an enrichment procedure that is independent of the function of the target gene (*17*). This procedure uses a positive selection for cells that have incorporated the targeting vector anywhere in the ES cell genome and a negative selection against cells that have randomly integrated the vector. The net effect is to enrich for cells containing the desired targeted mutation. The vector contains 10 to 15 kb of DNA homologous to the target gene, a *neo* inserted, along with its own strong promoter, into an exon of that sequence; and thymidine kinase gene from herpes simplex virus (HSV-tk) adjacent to the region of homology.

This vector was designed so that homologous recombination will results in insertion of the entire vector into the endogenous gene. This produces a duplication of a portion of *hprt*.

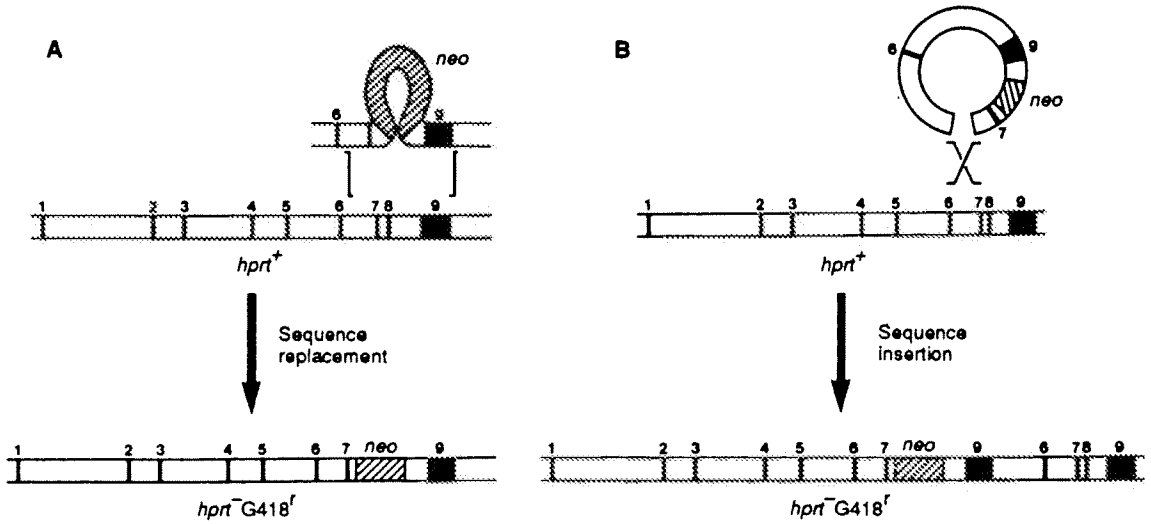

Fig. 2. Disruption of *hprt* by gene targeting with (**A**) a sequence replacement targeting vector or (**B**) a sequence insertion targeting vector. Vectors of both classes contain *hprt* sequences interrupted in the eighth exon with *neo* gene. With the sequence replacement vector, after homologous pairing between the vector and genomic sequences, a recombination event replaces the genomic sequence with vector sequences containing *neo*. Sequence insertion vectors are designed such that the ends of the linearized vector lie adjacent to one another on the *hprt* map. Pairing of such a vector with its genomic homolog, followed by recombination at the double-strand break,

result in the HSV-tk gene *not* being transferred into the target locus. The HSV-tk gene is lost during the process of homologous recombination because it is located distal to the region of homology between the vector and the target. Only cells in which random integration of the targeting vector has occurred will be able to retain the HSV-tk gene. This is predicted from the observation that most random insertions of exogenous, linearized DNA into the genome occur via their ends (*18*). Therefore, by using G418 to select *for* cells that contain a functional neo^r gene and by using gancyclovir to select *against* cells that contain a functional HSV-tk gene, we can enrich for cells in which the targeting event has occurred.

We have tested the above PNS approach by using it to disrupt the mouse *hprt*, *int-2*, *hox 1.2*, and *hox 1.3* genes (*17, 19*) and observed as much as a 2000-fold enrichment for targeted disruptions.

Targeting into Genes Not Expressed in ES Cells

An ideal strategy for disrupting a gene not expressed in ES cells has not yet emerged. Preliminary attempts in which the PNS procedure was used to mutate one such gene, the proto-oncogene *int-1*, suggests that the targeting frequency is much lower than that obtained at the *hprt*, *int-2*, *hox 1.2*, and *hox 1.3* loci (*20*). It may be possible to increase the enrichment factor by strategies designed to enhance the stringency of either the positive or negative component of the selection. As an example of the latter, to reduce the frequency of HSV-tk loss during transfection, two HSV-tk genes could be inserted into the targeting vector, one at each end, and then at least one should survive transfection to permit selection against cells containing random integrations. Placement of a large block of nonhomology at both ends of the linearized targeting vector, as is required to construct a double tk vector, does not affect the frequency of homologous recombination (*20*).

Among alternative approaches to consider are the use of pure screening procedures or enrichment procedures. Embedded within the decision of whether to use nonselective or selective protocols is the choice of whether to deliver the targeting vector to the recipient cells by microinjection or electroporation.

Microinjection into the nucleus results in very high frequencies of stable transfectants (10 to 20% of the cells receiving DNA) (*21*). However, only a single cell at a time can be microinjected, whereas electroporation allows large numbers of cells to be simultaneously transfected. After electroporation ~1% of ES cells are stable transfectants (*17*). The choice of transfection protocol should be determined, in part, by the method chosen to identify and isolate the ES cells in which the desired targeting event has occurred. Methods that rely solely on screening are dependent on the absolute targeting frequency and therefore microinjection should be considered. On the other hand, since enrichment procedures rely on selection they are not as dependent on the absolute targeting frequency and can take advantage of mass transfection protocols.

As already discussed, Zimmer and Gruss (*14*) used microinjection to deliver the targeting vector and reported that 1 in 150 cells receiving an injection of an altered *hox 1.1* genomic fragment yielded a disrupted *hox 1.1* gene. The reasons for this very high frequency of homologous recombination are not clear, but some or all of the following factors may have played a role: (i) delivery of the DNA by microinjection, (ii) absence of a selection protocol to isolate the ES cells in which a targeting event had occurred, (iii) disruption of the *hox 1.1* genomic fragment in the targeting vector with only a small (20-bp) insert, or (iv) a hot spot for recombination at the *hox 1.1* locus in ES cells.

Intrachromosomal gene conversion experiments in cultured mammalian cells suggest that the frequency of recombination events is inversely related to the length of nonhomology that must be corrected (*22*). Therefore, small disruptions in the targeting vector may favor higher targeting frequencies. This parameter has not, however, been directly measured in gene targeting experiments. Since the DNA substrates participating in intrachromosomal and gene targeting reactions may be quite different, extrapolating the results from one set of experiments to the other could be misleading.

On the basis of the above discussion, a number of alternative approaches can be suggested that may allow disruption of a gene not expressed in the recipient cell. Among these alternatives are: (i) microinjection of the targeting vector into the recipient cells followed by PCR to screen for the transfer of a small oligonucleotide insert to the desired target gene, (ii) electroporation of the targeting vector followed by PCR or (iii) electroporation of the targeting vector followed by an enrichment procedure, such as PNS, to select cell lines in which the targeting event has occurred. Each approach is associated with its own risks. If the targeting frequency with microinjection is in the range of one event per 10^4 cells, then this procedure is extremely labor-intensive. The absolute frequency of targeting events by the second procedure could be less than one event per 10^5 electroporated cells. Under these conditions, it becomes a challenge to use PCR to identify the targeting event. Finally, concerning the third approach, it is possible that the chromosomal environment surrounding a silent gene in ES cells may repress expression of the selectable marker. Under these conditions, enrichment procedures such as the described PNS procedure cannot be used.

The Future of Gene Targeting

As different methods of identifying successful targeting events are applied to more genes, their strengths and limitations will become evident. Current techniques should permit the generation of mouse mutants in many, if not all, genes. In particular, if the gene is expressed in ES cells, then enrichment procedures should allow the desired, targeted-modified ES cells to be found among a small number of selected clones (1 to 20). If, on the other hand,

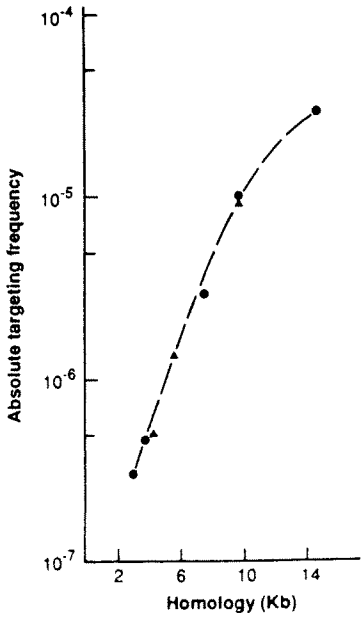

Fig. 3. The targeting frequency at the *hprt* locus as a function of the extent of homology between the targeting vector and the endogenous target. A series of replacement, Δ, and insertion, O, *hprt* targeting vectors were constructed that varied in the extent of homology to the endogenous *hprt*. *Hprt* sequences in each vector were interrupted in the eighth exon with *neo*. The amount of *hprt* sequence 3′ to *neo* was kept constant and the amount of *hprt* sequence 5′ to *neo* varied. The conditions for electroporation and selection for G418^r–6-TG^r ES cells were as described (*7*). All G418^r–6-TG^r cell lines were *hprt*- as a result of targeted disruption of *hprt*. Independence of the targeting events was guaranteed by only scoring G418^r–6-TG^r colonies that arose on separate plates. The absolute targeting frequency number of independent targeting events per total ES cells electroporated is plotted on a logarithmic scale as a function of the number of kilobases of *hprt* sequence contained within the targeting vectors.

expression of the gene of interest is not detectable in ES cells, the experimental options are less well defined.

With the above proviso, the future of gene targeting in mammalian systems appears very bright. Generation of specific mouse mutations via gene targeting should have a major impact on all phases of mammalian biology, including development, cancer, immunology, neurobiology, and human medicine. For example, recent molecular genetic analysis of development in *Drosophila* has revealed a network of genes that control the formation of its metameric pattern (*23*). On the basis of DNA sequence similarity, related genes, such as the *hox* genes, have been identified in the mouse (*24*). The embryonic expression patterns of these genes imply roles in establishing positional information during development. How closely the function of the mouse genes may parallel the function of the *Drosophila* homologs, if at all, remains to be determined. Targeted disruption of these genes may not only reveal the phenotypes associated with the inactivation of the individual genes, but, through epistasis and molecular analyses, may also help define the developmental network controlling early mouse morphogenesis.

Molecular analysis of tumors and transformed cells has revealed a plethora of genes contributing to malignant growth. The normal function of these proto-oncogenes is currently deduced from such features as their protein products, their cellular compartmentalization, and their expression pattern. Genetic dissection of their function in the mouse or in tissue culture systems should permit a more precise definition of the normal function of these genes, and potentially, a better appreciation of the role of oncogenes in causing malignancy. The development of genetically engineered mice in which the effects of deficiencies in anti-oncogenes, such as the retinoblastoma gene, could be studied would prove of great experimental value. On the other hand, the genetic function of some of the more ubiquitously expressed proto-oncogenes, such as *myc*, *fos*, and *jun*, may be more clearly defined in tissue culture systems rather than in the intact animal. ES cells could be a particularly attractive cell culture line for such an analysis. These cells are euploid and stable with respect to karyotype. Further, they can be induced to differentiate in vitro into many different cell types. If these factors participate in making early cell lineage decisions, then loss of function alleles could lead to a restriction in the spectrum of final differentiated cell types.

Immunology appears to be a particularly fertile field for genetic analysis by gene targeting. It is by far the best characterized cellular system in mammals, and extensive molecular analysis has identified many of the genetic components responsible for immunological diversity. Further, since the immune system is dispensable, analysis of many of the null phenotypes should be simplified.

Approximately 3500 different human genetic diseases are known. As the genes responsible for these diseases are identified and cloned, disruption of the corresponding genes in the mouse should provide useful models for these human diseases. Such models will facilitate analysis of the pathology of the disease and provide a system for the exploration of new therapeutic protocols including gene therapy.

A new arena for biotechnology is the application of transgenesis to the improvement of domestic animals and plants, as well as for the production of rare products such as pharmaceuticals within domestic animals (as discussed in this issue). These efforts will certainly be complemented and augmented by the use of gene targeting to modify the host genome. Transgenesis and gene targeting are often directed towards different ends, the former being used to gain new functions, the latter being used to augment or to generate loss of existing functions. As regulatory loops become better defined, it should become possible to alter such loops by gene targeting, thereby reducing the production of undesired products (such as fat content in meat) or increasing the production of desired products (such as pharmaceuticals in milk). The power of gene targeting resides in the ability of the experimenter to precisely choose both the gene to be modified and the specific change to be introduced.

REFERENCES AND NOTES

1. A. Gossler, T. Doetschman, R. Korn, E. Serfling, R. Kemler, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 9065 (1986); E. Robertson, A. Bradley, M. Kuehn, M. Evans, *Nature* **323**, 445 (1986).
2. M. Hooper, K. Hardy, A. Handyside, S. Hunter, M. Monk, *Nature* **326**, 292 (1987); M. R. Keuhn, A. Bradley, E. Robertson, M. Evans, *ibid.*, p. 295.
3. S. Thompson, A. R. Clarke, A. M. Pow, M. Hooper, D. W. Melton, *Cell* **56**, 313 (1988).
4. M. Evans and M. Kaufman, *Nature* **292**, 154 (1981); G. Martin, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 7634 (1981); T. Doetschman, F. Williams, N. Maeda, *Dev. Biol.* **127**, 224 (1988).
5. J. Paszkowski, M. Baur, A. Bogucki, I. Potrykus, *EMBO J.* **7**, 4021 (1988).
6. K. R. Folger, K. R. Thomas, M. R. Capecchi, *Cold Spring Harbor Symp. Quant. Biol.* **49**, 123 (1984); F. L. Lin, K. Sperle, N. Sternberg, *ibid.*, p. 139; O. Smithies, M. A. Koralewski, K. -Y. Song, R. S. Kurcherlepati, *ibid.*, p. 161; A. J. H. Smith and P. Berg, *ibid.*, p. 171.
7. M. R. Capecchi, *Trends Genet.* **5**, 70 (1989).
8. K. R. Thomas and M. R. Capecchi, *Cell* **51**, 503 (1987).
9. T. Doetschman, R. G. Gregg, N. Meada, M. Hooper, D. W. Thompson, O. Smithies, *Nature* **330**, 576 (1987).
10. T. Doetschman, N. Meada, O. Smithies, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8553 (1988).
11. K. B. Mullis and F. Faloyna, *Methods Enzymol.* **155**, 335 (1987).
12. H. S. Kim and O. Smithies, *Nucleic Acids Res.* **16**, 8887 (1988).
13. A. L. Joyner, W. C. Skarnes, J. Rossant, *Nature* **338**, 153 (1989).
14. A. Zimmer and P. Gruss, *ibid.*, p. 150.
15. M. Jasin and P. Berg, *Genes Dev.* **2**, 1353 (1988).
16. J. M. Sedivy and P. A. Sharp, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 227 (1989).
17. S. L. Mansour, K. R. Thomas, M. R. Capecchi, *Nature* **336**, 348 (1988).
18. K. R. Folger, E. A. Wong, G. Wahl, M. R. Capecchi, *Mol. Cell. Biol.* **2**, 1372 (1982); D. B. Roth, T. N. Porter, J. M. Wilson, *ibid.* **5**, 2599 (1985); K. R. Thomas, K. R. Folger and M. R. Capecchi, *Cell* **44**, 419 (1986).
19. D. Kostic and M. R. Capecchi, unpublished results.
20. K. R. Thomas and M. R. Capecchi, unpublished results.
21. M. R. Capecchi, *Cell* **22**, 479 (1980).
22. A. Letsue and R. M. Liskay, *Genetics* **117**, 759 (1987).
23. For a recent review see P. W. Ingham, *Nature* **335**, 25 (1988).
24. For recent reviews see P. W. M. Holland and B. L. M. Hogan, *Genes Dev.* **2**, 773 (1988); G. R. Dressler and P. Gruss, *Trends Genet.* **4**, 214 (1988).